Mining User Behaviours: A Study of Check-in Patterns in Location Based Social Networks

The University Of Sheffield.

Daniel Preoțiuc-Pietro
daniel@dcs.shef.ac.uk

Trevor Cohn
t.cohn@dcs.shef.ac.uk

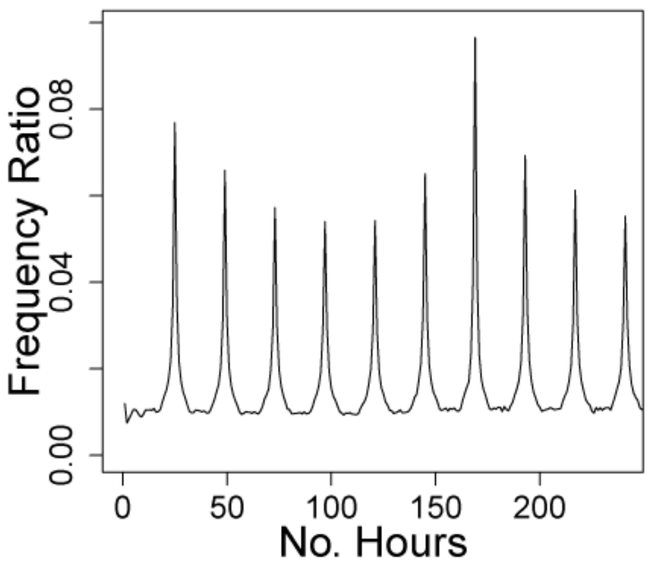


Fig. 1 Returning frequency ratio

MOTIVATION

- understanding the patterns underlying human mobility is of an essential importance to a range of applications (e.g. recommender systems)
- previous research limited by lack of data & privacy
- LBSNs offer us data and metadata about user movements

DATASET

- consists of `frequent' users of foursquare
- novel data collection method focused on users extracted from September 2011 via Twitter

No. users	No. checkins	No. checkins/user
9167	959,122	104.6 ± 49.4

DATASET PROPRIETIES

1 Time Distribution
- regular patterns for each day of the week: weekdays with similar shapes and week-ends more evenly distributed

2 Interevent times & distances
- the frequencies between consecutive check-ins by a user decrease as the time frame increases, with a `bump' at 8-10h
- the distance between consecutive check-ins follows a power law x^b with b=-1.56
- frequent users make small moves and rarely larger moves

3 Returning probability
Fig.1 - the probability of a user returning to the same place after h hours after being observed h hours ago
- we observe strong daily and weekly patterns
- it's more likely for a person to do the same thing as he did the same time last week, than the same time yesterday

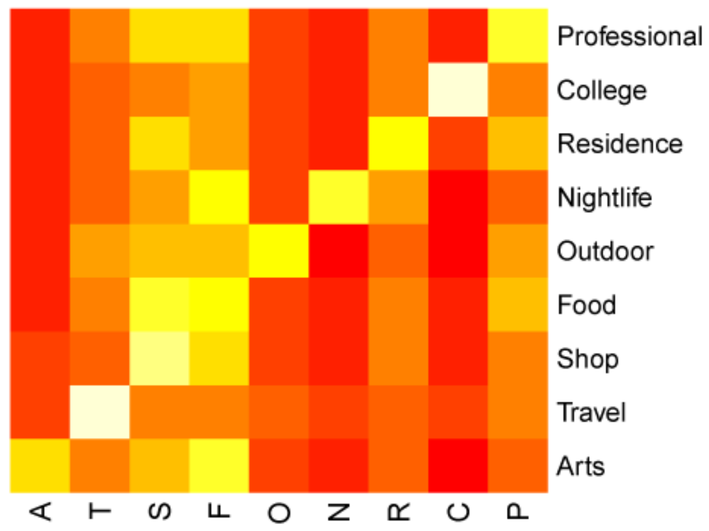


Fig. 2 Transition frequency between venue types

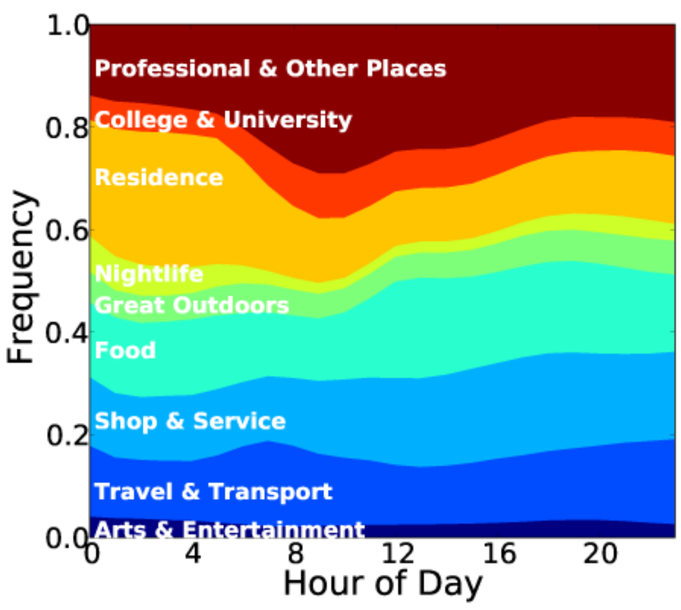
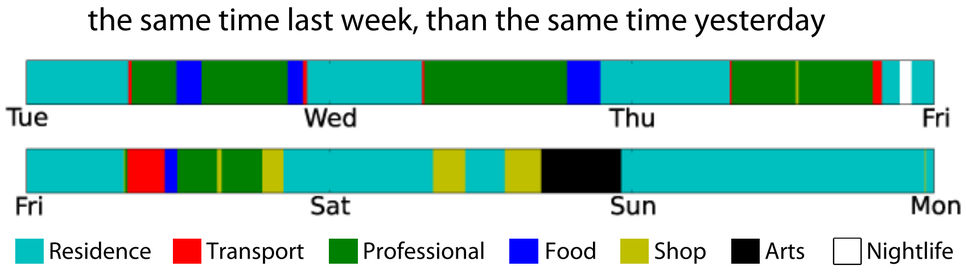


Fig. 3 Weekday location distribution



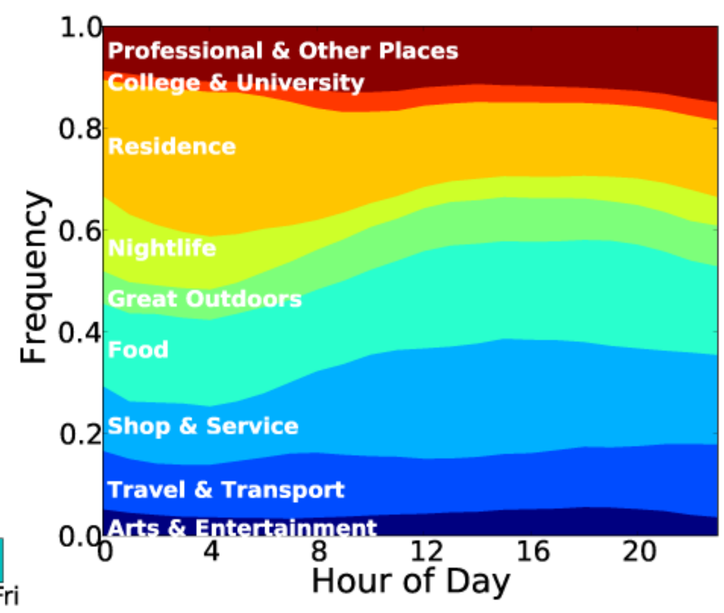


Fig. 4 Saturday location distribution

EXPLORING VENUE INFORMATION

1 Check-ins and categories
- for every venue type (9 types), there is a different distribution, but all show regularities in periodic behaviour (e.g. `Professional' is more frequent on weekdays with peaks in the morning and after lunch, 'Nightlife' is more frequent in the weekend evenings)

2 Daily patterns
- we study both check-in and present location of users (considered at the place of the last check-in)

Fig.3,4 - distribution of locations of users on weekdays and Saturdays

3 Interevent times
- time distributions vary based on the venue type (e.g. `Transport' has usually very short times, `Residence' is more smoothly distributed)

FUTURE WORK

- analysing other venue information (photos, tags, tips, etc.)
- performing regional analysis
- models of predicting future movements that incorporate periodicities in the data

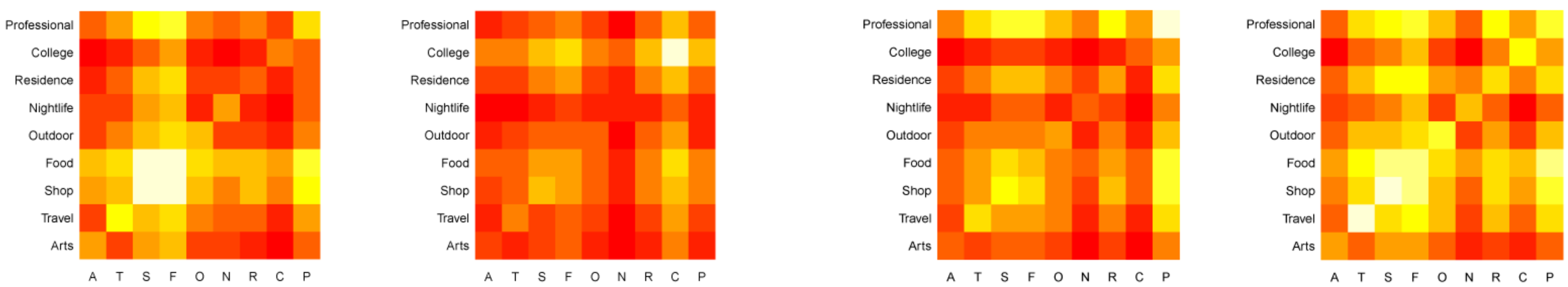
ANALYSING HUMAN MOVEMENT

1 Transitions
Fig.2 - there are frequent transitions between venues of similar categories
- the heatmap is not symmetric, highlighting some predictible behaviors

2 Behavioural clustering of users
- k-means clustering (k=8) on the user's transition matrices
- Fig.5 - some centroids of the clusters, showing `prototype behaviours' for each category of users

3 Predicting future user movements
- Markov Models are very popular and have good results
- in LBSNs prediction is poor and decays with gathering more data because of venue sparsity
- results show that incorporating temporal patterns, we obtain improvements over baseline methods

Method	Accuracy
Random Baseline	11.11%
Most Frequent Category (Markov-0)	35.21%
Markov-1 (with backoff to Markov-0)	36.13%
Markov-2 (with backoff to Markov-1)	34.21%
Most Frequent Hour	38.92%
Most Frequent Day of Week and Hour	40.65%



Cluster 7 (1282 users) `Shoppers'
Cluster 8 (115 users) `Students'
Cluster 3 (630 users) `Businessmen'
Cluster 4 (4646 users) `Generic user'

Fig. 5 Centroids of k-means clustering on user's venue category transition matrices